

A Simple Modularity Measure for Search Spaces based on Information Theory

Peter Dauscher¹, Daniel Polani², Richard A. Watson³

¹ Johannes Gutenberg University Mainz, Germany

² University of Hertfordshire, Hatfield, UK

³ University of Southampton, Southampton, UK
dauscher@informatik.uni-mainz.de

Abstract

Within the context of Artificial Life the question about the role of modularity has turned out to be crucial, especially with regard to the problem of evolvability. In order to be able to observe the development of modular structure, appropriate modularity measures are important. We introduce a continuous measure based on information theory which can characterize the coupling among subsystems in a search problem. In order to illustrate the concepts developed, they are applied to a very simple and intuitive set of combinatorial problems similar to scenarios used in the seminal work by Simon (1969). It is shown that this measure is closely related to the classification of search problems in terms of *Separability*, *Non-Decomposability* and *Modular Interdependency* as introduced in (Watson and Pollack, 2005).

Introduction

Central to the studies of Artificial Life is the understanding of how complex systems arise from simpler ones.

One of the particular challenges facing the field is that there are no canonical restrictions or a priori limitations of the models used; this is unlike fields, such as, e.g., physics or chemistry, where there are universal laws of conservation of energy, mass and other quantities.

In Artificial Life research, the lack of such natural constraints gives rise to a vast spectrum of computational models and simulated physical worlds which are not restrained by natural limitations. While, on the one hand, this is an advantage as it allows to study "worlds as they could be" as opposed to "world as they are", on the other hand its arbitrariness also creates problems. The different models do not need to share any constraints, and thus statements that can be generalized from one model to another are difficult to make. In addition, this situation tends to favour research approaches where the development and study of models is driven by phenomenological considerations. In view of this situation, it has been felt for some time that a better principled understanding of such systems is necessary: this would provide a systematic approach to analyse, predict and construct complex Alife systems.

To achieve this, a consistent mathematical language and formalism describing these systems is necessary. It becomes

increasingly clear that information theory, via its wide array of ramifications into different fields and its powerful mathematical arsenal is one of the most promising candidates. The last years have seen a surge in highly successful applications of information theory to complex and Artificial Life systems (Adami, 1997; Tishby et al., 1999; Shalizi and Crutchfield, 2002; Klyubin et al., 2005). Information theory both provides tools for the analysis and for the construction of complex systems.

In order to answer important Artificial Life related questions using information theory, however, we have to translate the respective model as well as the question itself into the right language. The present paper does that with the most relevant question of modularity of a search function. We will find that even in the simple case discussed here, this is a non-obvious task. This is also the reason that, rather than to design an elaborate simulation model, here we concentrate on the "essence" of aforementioned translation process. Instead of justifying the result of this translation on the phenomenological level of an intricate simulation, we will prefer to concentrate on a very minimalistic but well known scenario that, so we hope, will provide the relevant insights into the translation process in a more transparent way.

We also expect that explaining the concepts using intuitive examples without distracting additional complexity will provide the readers a reliable basis for the decision whether to use this measure for their special purposes or not.

Modularity

In the studies of complex systems and Artificial Life, a central question is how it is possible that, over time, systems can emerge with ever increasing complexity. This question is particularly prominent if one considers the Darwinian evolution which, from a naive point of view, appears to be mainly a *Random Heuristic Search* (Vose, 1999) with a large test population which is performed over very large timescales. However, one may doubt whether the massive parallelism and the large timescales alone can explain the enormous complexity of living organisms that we observe.

The hierarchical structures in living systems, reaching

from the molecular level up to whole organisms and – even further – to complex societies, might remind one to methods, that intelligent designers use, especially in the design of large software systems. In the latter, with the advent of the software crisis in the 1970s (Zuser et al., 2001), it became clear that large monolithic software systems in which each part depends on many others (a form of nonlocality) are practically unmanageable. Even if they should work reasonably reliably at a certain point in time, they cannot easily be adapted to new tasks. This is being solved by introducing *modules* which solve subproblems independently from the rest of the system and organizing larger systems by building them up from these smaller, manageable modules.

Adaptability, in turn, is one of the central motifs of natural evolution. Therefore, the question arises whether evolution manages complexity in a similar way as human software engineers, via modularity. It turns out that there are several phenomena in nature that can be construed as exhibiting elements of modularity. The most prominent examples are the recombination of distinct chromosomes in sexual reproduction and the crossover operator (which is construed by researchers of artificial evolution as to be preserving *building blocks* which encode for separable, i.e. modular, properties of the phenotype).

Therefore, modularity has become a central issue in the study of evolved biological functionality (Snel and Huynen, 2004) and it appears quite natural that there have been many approaches to characterize or to measure modularity or to investigate how more and more modular structures emerge in the course of evolution processes (Wagner, 1995; Calabretta et al., 1998; Dauscher and Uthmann, 2002; Dauscher and Uthmann, 2005).

Information theory is being used as a nonlinear correlation measure to discover suspected modules in genomes (Steuer et al., 2002). Recent approaches have also shown that information theory can be used to define a measure for the modularity of networks (Ziv et al., 2005; Kashtan and Alon, 2005; Hallinan, 2004), neural structures (Hüsken et al., 2001) or of simple dynamical systems (Polani et al., 2005). Also the modularity of evolutionary search has already been tackled using information theoretic concepts (Mühlenbein and Höns, 2005). It is to be noted, however, that this approach requires a quite intricate mathematical formalism.

In the paper, we will develop a different, comparatively simple and intuitive approach to characterize modularity of search problems.

Three Simple Scenarios

In the pioneering work “The Sciences of the Artificial” (Simon, 1969), one of the most prominent precursors of the Artificial Life research field, several types of imaginary safe locks are used to illustrate the concept of subsystems that are “nearly decomposable”. Note that these examples are

not as arbitrarily chosen as it might appear at first sight. On the contrary, they can be conceived as a striking analogue to difficult search problems, where the goal is to find solutions better than all solutions found, yet. These problems are closely related to the concept of Evolvability (Altenberg, 1994), which, in turn, tackles one of the most important questions of Artificial Life: “How can structures of extreme complexity emerge from evolutionary processes?”.

In order to illustrate the concepts developed in this paper we introduce three very simple scenarios each dealing with a safe-lock consisting of 10 binary switches. This corresponds to $2^{10} = 1024$ possible combinations; we assume that only one of these combinations really opens the safe. We consider a safe cracker having no a priori information about the right combination.

The difference in the three scenarios lies in the way the locks reveal useful information about the opening combination: we distinguish the *Silent Lock*, the *Revealing Lock* and the *Ambiguously Revealing Lock*.

The *Silent Lock* is a safe-lock as it is intended: it does not reveal anything about the right combination until it has been found. In the worst case a safe cracker will have to try out 1023 combinations before knowing the correct combination.

In the *Revealing Lock* the safe-lock makes a clicking sound in the case of the first 5 bits correctly set and a (different) clicking sound if the other 5 bits are set correctly. This makes things quite easier for a safe-cracker, since he can start by determining the first 5 bits and the last 5 bits separately. In the worst case, this procedure will take $31+31=62$ trials. Note that these considerations assume that the safe cracker knows the facts described above, which also includes the knowledge about which 5 bits are the “first” 5 bits. Without this knowledge, the clicking sounds produced might be a help but on a quite smaller scale.

The *Ambiguously Revealing Lock* also makes clicking sounds, however not only for the right combination of the respective 5-bit part but also for a different combination which, however, is not at all useful to open the lock. We assume it to be impossible to distinguish the sounds of the right from the wrong combinations.

One can imagine that the ambiguity makes it more complicated again to open the safe. A possible way to find out the opening combination is to determine the two clicking partial solutions for each part. Then, there are four possible combinations of these partial solutions. In the worst case, this procedure will take $31+31+3=65$ trials.

Non-Decomposability, Separability and Modular Interdependency

(Watson and Pollack, 2005), following Simon (1969), discusses a number of hypothetical locks like those above to illustrate a combinatorial search problem that can be divided into sub-problems. However, the intent in that work was not just to illustrate modules that are entirely indepen-

dent (as in the revealing lock) but also a case where useful modules are available to reduce the combinatorics of the search involved without assuming that the modules are entirely separable - this is the purpose of the ambiguously revealing lock example. This form of incomplete modularity draws from Simon's intuition of 'nearly decomposable' systems (1969) but, although Simon focused on this concept extensively, his lock example and other examples (including the watchmaker's parable) did not provide a combinatorial search problem that was nearly decomposable in a wholly satisfying way (Watson 2005).

Although the idea of nearly decomposable systems is intuitive, this kind of 'modular but not completely modular' notion becomes problematic, and does not capture all of what we might hope to capture when applied to combinatorial search. The issue is as follows: if we are able to determine the correct setting for a subset of the variables in a combinatorial search problem without regard to the setting of other variables then the modules are completely independent. If the correct setting for a subset of variables is a little bit different depending on the setting of other variables in the system then this seems like a reasonable approximation to partial modularity. But, in this kind of near-decomposability, it is difficult to make strongly significant interdependencies between modules without removing the utility of the modules altogether. However the assumption that inter-module dependencies must be weak is not necessary and limits the utility of hierarchical modularity (Watson 2005).

In contrast to this simplistic notion, note that in the ambiguously revealing lock we can see that the two clicking settings might have no bits in common - one could be 00000 and the other could be 11111 - but this is not important to the combinatorial reduction the clicks provide. We also see that if the safe is set to the 'deceptive' clicking position, this results in complete lock-out - the ability to open the safe is utterly dependent on having the correct setting in both modules, so the inter-module dependency in this problem could not be stronger. Nonetheless, the fact that the clicking positions can be identified means that the modules are highly effective in revealing the correct setting to open the safe. Thus the modules in this problem are very clear and significant in the way they reduce the combinatorics of the problem, but this is not opposed to the fact that resolving the dependencies between the modules is critical.

(Watson and Pollack, 2005) suggests a coarse grained classification of the interaction in an optimization problem like this. In this paper the term 'decomposable' is defined by comparing the number of settings that a module can take, C , with the number of different settings, C' , of that module that are optimal *for some context*. That is, in the lock example, if the clicks correspond to a suboptimal solution, we can identify these as configurations that can be saved and re-tried with combinations of the other module. If this is true for both modules then we can greatly reduce the number of

total configurations that need to be tested. Even though we cannot uniquely identify which settings are globally optimal (actually open the safe) independently of the other module, i.e. $C' > 1$, we can reduce the number of settings we need to consider from $C = 32$ down to $C' = 2$. Watson defines a 'decomposable' system as one where $C' < C$, a 'separable' system as one where $C' = 1$, and suggests that cases of particular interest are decomposable but not separable systems, i.e., $1 < C' < C$, which he calls "Modular Interdependency". By these definitions, the silent lock example is non-decomposable, the revealing lock example is separable, and the ambiguously revealing lock example exhibits Modular Interdependency.

These terms are conceptually useful and feed into the measure of modularity we define in this paper. However they also have some conflicts with our basic intuitions about modularity in disconnected systems and other cases (Polani et al., 2005). In this paper we convert this classification into a more rigorous information theoretic measure.

Formalization within the Framework of Information Theory

Let our system S (the safe lock) be subdivided into two subsystems S_1 and S_2 (the first 5 bits and the other 5 bits). The fact that we do not know anything about the right combination before having made experiments is now expressed by the corresponding random variables S , S_1 and S_2 , each of which is equally distributed.

The first question we ask is: What is the difference between the three scenarios from the information-theoretic point of view? Therefore, we consider first the uncertainty about the right combination which is described by the entropy $H(S)$. It is easy to see that $H(S) = 10$ in all three of the scenarios. Obviously, this naive measure does not adequately represent the effort of the safe-cracker or the coupling of the subsystems.

This is due to the fact that the Shannon entropy (uncertainty) implicitly assumes that the system can be asked an *optimally selected set of questions*: for instance, in our scenario, the 10 bit of uncertainty could correspond to the ability to ask for each bit separately whether it is set to "1" or to "0" in the right combination. In this case, ten of these optimal binary questions would be sufficient in order to determine the complete right combination.

In constructing safe-locks, of course, engineers aim at quite the opposite: they limit the set of askable questions as to make it as difficult as possible to determine the right combination just by trial and error. What makes the *Silent Lock* particularly hard to crack is that in order to determine the right state of any subsystem you have to know the state of the whole rest: It is impossible to do ask useful questions considering, e.g., only 5 bits. The only kind of askable questions is "*Does combination ... open the safe-lock?*". Therefore, you have to check (in the worst case) $2^{10} - 1 = 1023$

combinations before definitely knowing the opening combination.

In the *Revealing Lock*, the set of usefully askable questions is larger: Again, one can ask whether a given combination opens the lock; in addition, one may also ask whether a partial solution of 5 bits produces a clicking sound (and thus is the right partial solution for these 5 bits). That means that one can determine the right partial solution without having to learn anything about the rest (and thus avoiding the effort of asking many questions).

In the *Ambiguously Revealing Lock* we find an intermediate case: The set of askable questions is the same, however the answer “click” to the question whether a partial solution of the first 5 bits is correct may be useful but is only partially informative.

We will now formalize the notions introduced above. Knowledge about the right combination can only be obtained by suitable experiments. We will therefore introduce the concept of “measurements” where a measurement \mathbf{M} is a tuple $(M^{(1)}, \dots, M^{(k)})$ of random variables $M^{(i)}$ each containing the outcome of a given experiment. An experiment may be the testing a combination of the first or last 5 bits (completely neglecting the respective rest) whether it produces a clicking sound. Another experiment could be to test a 10-bit combination whether or not it opens the safe-lock.

In the information-theoretic context, we treat the measurements like ordinary random variables by setting:

$$\begin{aligned} H(\mathbf{M}) &= H(M^{(1)}, M^{(2)}, \dots, M^{(k)}) \\ H(X|\mathbf{M}) &= H(X|M^{(1)}, M^{(2)}, \dots, M^{(k)}) \\ I(X;\mathbf{M}) &= I(X;M^{(1)}, M^{(2)}, \dots, M^{(k)}) \end{aligned}$$

A Measure Based on the Coupling of Modules

Let us ignore for the moment the possibility that there might be any correlation between the random variables S_1 and S_2 , i.e. $I(S_1; S_2) = 0$. We will discuss a more general version including $I(S_1; S_2) \geq 0$ in the next section. We now define a (preliminary) measure for coupling:

$$K = \min_{\{\mathbf{M}_1, \mathbf{M}_2 \mid H(S|\mathbf{M}_1, \mathbf{M}_2) = 0\}} [I(\mathbf{M}_1; S_2) + I(\mathbf{M}_2; S_1)] \quad (1)$$

That means: we consider two measurements $\mathbf{M}_1, \mathbf{M}_2$ that are together sufficient to identify the necessary information for the whole system. In one extreme case the measurements \mathbf{M}_1 and \mathbf{M}_2 can be applied to the corresponding subsystems S_1 and S_2 without the necessity to obtain information from the respectively other ones. Vice versa, any information which has necessarily to be obtained from the respectively other one is a measure for the coupling of the subsystems. It is easy to see that K has an upper bound, namely $\min[H(S_1), H(S_2)]$, corresponding to the opposite extreme

case. The boundedness can be shown as follows: let us assume $H(S_1) \geq H(S_2)$ without loss of generality. Then we consider a measure \mathbf{M}_1 obtaining all information about the whole system alone and \mathbf{M}_2 an “empty” measurement (containing no experiments at all). Then K cannot be larger than $H(S_2)$, the smallest of the two values $H(S_1)$ and $H(S_2)$.

The both extreme cases and the range between them can now be characterized in a quite natural way by three categories introduced above:

Separability	$K = 0$
Non-Decomposability	$K = \min[H(S_1), H(S_2)]$
Modular Inter-dependency	$0 < K < \min[H(S_1), H(S_2)]$

In order to depict these concept graphically, we introduce diagrams rather similar to the typical Venn-Diagrams that are frequently used to describe information-theoretic concepts, as, e.g., in (Adami, 1995); one needs to exert some care using them for inequality proofs, (Yeung et al., 2002). In the diagrams shown in Fig. 1, the measurements $\mathbf{M}_1, \mathbf{M}_2$ (more exactly: their entropy) are represented as boxes stacked upon each other. The entropies of the subsystems S_1 and S_2 , in contrast, are represented as boxes positioned side by side.

Application to the Example Scenarios

Let us review now our three locks before the theoretical background developed so far.

In the *Revealing Lock*, we can perform two measurements separately - as described above: one (without loss of generality \mathbf{M}_1) that only yields information about S_1 (by using only the clicking sound information) and one that only yields information about S_2 , using the slightly different clicking sound. Thus, $I(\mathbf{M}_1; S_2) = 0$ and $I(\mathbf{M}_2; S_1) = 0$ and this immediately leads to $K = 0$: the *Revealing Lock* is separable.

The *Silent Lock*, in contrast, is non-decomposable. This can be seen by noting that none of the single experiments that can be performed reveals more about one of the subsystems than about the other one. Hence, independently of how one distributes the necessary experiments onto the two measurements \mathbf{M}_1 and \mathbf{M}_2 ,

$$\begin{aligned} I(\mathbf{M}_1; S_2) &= I(\mathbf{M}_1; S_1) \\ I(\mathbf{M}_2; S_1) &= I(\mathbf{M}_2; S_2) \end{aligned} \quad (2)$$

It is easy to see that the requirement $H(S|\mathbf{M}_1, \mathbf{M}_2) = 0$ is equivalent to $I(S; \mathbf{M}_1, \mathbf{M}_2) = H(S)$. Since S_1 and S_2 have no mutual information, we find $H(S) = H(S_1) + H(S_2)$ and $I(S; \mathbf{M}_1, \mathbf{M}_2) = I(S_1; \mathbf{M}_1, \mathbf{M}_2) + I(S_2; \mathbf{M}_1, \mathbf{M}_2)$. We therefore can compute

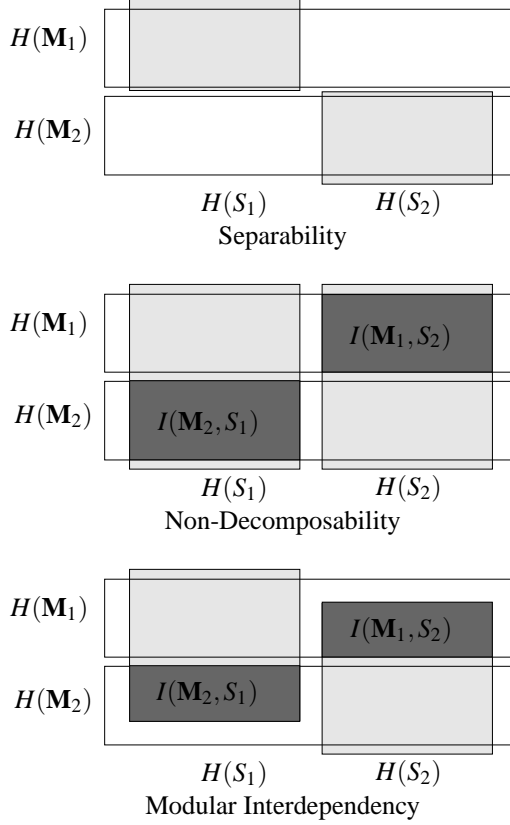


Figure 1: Graphical representation three categories *Separability*, *Non-Decomposability* and *Modular Interdependency*. The area colored in dark gray correspond to the measure K as introduced in this section

$$H(S) = I(S; \mathbf{M}_1, \mathbf{M}_2) \quad (3)$$

$$\leq I(S_1; \mathbf{M}_1) + I(S_1; \mathbf{M}_2) + I(S_2; \mathbf{M}_1) + I(S_2; \mathbf{M}_2) \quad (4)$$

$$\stackrel{(2)}{=} 2(I(S_1; \mathbf{M}_2) + I(S_2; \mathbf{M}_1)) \quad (5)$$

where “ \leq ” is “ $=$ ” in the case of vanishing mutual information between the two measurements. Hence, we can say that

$$K = \frac{1}{2}H(S) = 5 = \min(H(S_1), H(S_2)) \quad (6)$$

indicating that the *Silent Lock* is non-decomposable.

In the *Ambiguously Revealing Lock*, measurements about the clicking sounds alone will not reveal the unique right combination. Further experiments $M^{(i)}$, have to be performed to find the really opening one of the 4 remaining combinations (yielding 2 bits of information). Each of these experiments, however, reveals exactly as much about S_1 as about S_2 . Therefore, we can apply exactly the same argument as that of the silent lock above to the 4 remaining combinations. It follows that half of the two bits of information contributes to the term $[I(\mathbf{M}_1; S_2) + I(\mathbf{M}_2; S_1)]$: we find $K = 1$ corresponding to the case of Modular Interdependency.

Considering possible mutual information

So far, we have neglected the possibility that two subsystems S_1 and S_2 might not be independent of each other but that there is some mutual information between the corresponding random variables: $I(S_1, S_2) > 0$.

For our measure, this possibility has an important consequence: Let us consider a *Revealing Lock* similar to the one before with one additional property: We know that in the right overall combination the first two bits of the first 5-bit subsystem must have the same values as the first two bits of the other 5-bit subsystem. Each $M^{(i)}$ considering these first two bits necessarily reveals something about both subsystems ($K = 2$) Hence, this safe-lock cannot be separable any more. However, somebody who does not know this additional constraint will not see the difference to the normal Revealing Lock. Furthermore, this kind of coupling does not increase the necessary effort (as in the *Silent Lock*) but will – in contrast – even decrease it: Having found the right combination of, e.g. S_1 , only three bits from S_2 have to be determined.

So it is intuitive to require that for this lock again a reasonable measure K should have the value $K = 0$. We propose

$$K = \min_{\{\mathbf{M}_1, \mathbf{M}_2\} | H(S | \mathbf{M}_1, \mathbf{M}_2) = 0} [I(\mathbf{M}_1; S_2) + I(\mathbf{M}_2; S_1)] - I(S_1; S_2) \quad (7)$$

as an appropriate measure. The only difference between this measure and the one introduced before is the term $I(S_1, S_2)$

subtracted. It seems plausible that the newly introduced measure K cannot become negative; however, at this moment, we do not yet possess a formal proof.

Furthermore, also the criteria for a system to be separable, non-decomposable or modular interdependent, have to be adjusted.

Separability	$K = 0$
Non-Decomposability	$K = \min [H(S_2 S_1), H(S_2 S_1)]$
Modular Inter-dependency	$0 < K < \min [H(S_2 S_1), H(S_2 S_1)]$

The most complex case of decomposable but not separable systems is represented graphically in Fig. 2.

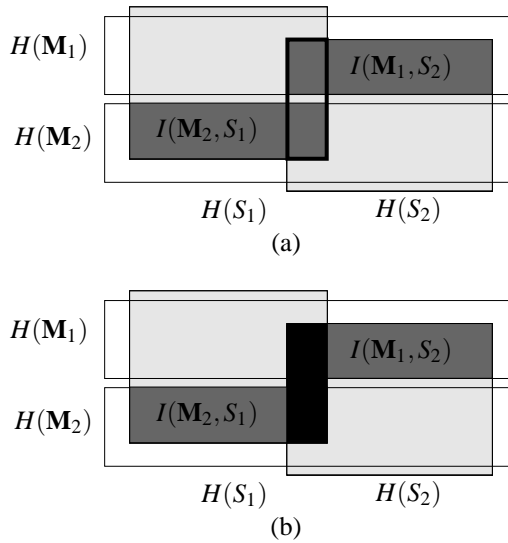


Figure 2: a) Graphical representation of the case of “Modular Interdependency” including mutual information ($I(S_1;S_2) > 0$). The mutual information is framed with a thick line in the above figure. b) In the measure, the mutual information $I(S_1;S_2)$ is subtracted (black box): the measure corresponds to the gray areas.

Note that in the case of $I(S_1;S_2) = 0$ the formula for K as well as the conditions equal exactly the measure introduced before.

Conclusion and Outlook

Starting from the concepts of *Separability*, *Non-Decomposability* and *Modular Interdependency* in (Watson and Pollack, 2005) we have developed a measure based on information theory which characterizes the coupling of modules in a more general and more fine-grained and thus, we believe, useful way. Because of the importance of search problems in general and especially modularity and evolvability, it appears more than likely that the measure introduced here can serve as a useful research tool, especially in the Complex Systems and Artificial Life context.

The generality of information theoretic approaches can be considered as one of the major advantages of our concepts.

After having investigated the properties of the measures in the simple scenarios considered here, it is, of course, appropriate to apply it to more typical Artificial Life scenarios. A first step we think of is to consider Simple Genetic Algorithms having a well defined fitness landscape (as, e.g. Royal Road Fitness Landscapes or Terraced Labyrinth Landscapes (Crutchfield and van Nimwegen, 2001) or Hierarchical If-and-only-if (HIFF) models (Watson and Pollack, 2005)). It could also be interesting to know whether there is a relation between the measure introduced here and, e.g. the ability to find a good solution in a given amount of time or generations. Furthermore, it seems promising to look for correlations with measures of evolvability as defined in (Altenberg, 1994)) in concrete simulation scenarios, which would underpin the assumption that modularity and evolvability are closely related, indeed. A next, quite more ambitious step would be to look for a connection of the modularity of a search space (as introduced here) and the modularity of the emerging individuals using measures as introduced in (Polani et al., 2005) or (Dauscher and Uthmann, 2005).

References

- Adami, C. (1995). On Modelling Life. *Artificial Life*, 1:429–438.
- Adami, N. (1997). *Introduction to Artificial Life*. Springer.
- Altenberg, L. (1994). The evolution of evolvability in genetic programming. In Kinnear, J. K. E., editor, *Advances in Genetic Programming*. MIT Press.
- Calabretta, R., Nolfi, S., Parisi, D., and Wagner, G. P. (1998). A case study of the evolution of modularity: towards a bridge between evolutionary biology, artificial life, neuro- and cognitive science. In Adami, C., Belew, R., Kitano, H., and Taylor, C., editors, *Proc. of 'Artificial Life VI', Los Angeles, June 26-29*. MIT Press.
- Crutchfield, J. and van Nimwegen, E. (2001). The evolutionary unfolding of complexity. In *Proc. Evolution as Computation, DIMACS workshop, 1999*. Springer, New York.
- Dauscher, P. and Uthmann, T. (2002). On self-organizing modularization of individuals in evolutionary scenarios. In Polani, D., Kim, J., and Martinetz, T., editors, *Proc. Fifth German Workshop on Artificial Life, March 18-20, 2002, Lübeck, Germany*. Akademische Verlagsgesellschaft Aka GmbH, Berlin.
- Dauscher, P. and Uthmann, T. (2005). Self-organized modularity in evolutionary algorithms. *Evolutionary Computation*, 13(3):303–328.

- Hallinan, J. (2004). Gene duplication and hierarchical modularity in intracellular interaction networks. *Biosystems*, 74(1-3):51–62.
- Hüsken, M., Igel, C., and Toussaint, M. (2001). Task-dependent evolution of modularity in neural networks – a quantitative case study. In *2001 Genetic and Evolutionary Computation Conference (GECCO 2001) – Late-Breaking Papers*, pages 187–193.
- Kashtan, N. and Alon, U. (2005). Spontaneous evolution of modularity and network motifs. *Proceedings of the National Academy of Science*, 102(39):13773–13778.
- Klyubin, A. S., Polani, D., and Nehaniv, C. L. (2005). Empowerment: A universal agent-centric measure of control. In *Proceedings of the 2005 IEEE Congress on Evolutionary Computation*, volume 1, pages 128–135. IEEE Press.
- Mühlenbein, H. and Höns, R. (2005). The Estimation of Distributions and the Minimum Relative Entropy Principle. *Evolutionary Computation*, to appear.
- Polani, D., Dauscher, P., and Uthmann, T. (2005). On a quantitative measure for modularity based on information theory. In *Proc. of the VIIIth European Conference on Artificial Life (ECAL 2005)*. Springer.
- Shalizi, C. R. and Crutchfield, J. P. (2002). Information bottlenecks, causal states, and statistical relevance bases: How to represent relevant information in memoryless transduction. *Advances in Complex Systems*, 5(1):91–95.
- Simon, H. (1969). *The Sciences of the Artificial*. MIT Press.
- Snel, B. and Huynen, M. A. (2004). Quantifying modularity in the evolution of biomolecular systems. *Genome Research*, 14:391–397.
- Steuer, R., Kurths, J., Daub, C. O., Weise, J., and Selbing, J. (2002). The mutual information: detecting and evaluating dependencies between variables. *Bioinformatics*, 18:231–240, Suppl. 2.
- Tishby, N., Pereira, F. C., and Bialek, W. (1999). The information bottleneck method. In *Proceedings of the 37th Annual Allerton Conference on Communication, Control, and Computing*, pages 368–377.
- Vose, M. D. (1999). *The Simple Genetic Algorithm*. MIT Press.
- Wagner, G. P. (1995). Adaptation and the Modular Design of Organisms. In Morán, F. et al., editors, *Advances in Artificial Life*, pages 317–328. Springer Verlag, Berlin.
- Watson, R. and Pollack, J. (2005). Modular interdependency in complex dynamical systems. *Artificial Life*, 11(4):445–457.
- Yeung, R. W., Lee, T. T., and Ye, Z. (2002). Informationtheoretic characterizations of conditional mutual independence and markov random fields. *IEEE Transactions on Information Theory*, 48(7):1996–2011.
- Ziv, E., Middendorf, M., and Wiggins, C. (2005). An information-theoretic approach to network modularity. *Physical Review E*, In press. arXiv:q-bio.QM/0411033v1.
- Zuser, W., Biffl, S., Grechenig, T., and Köhle, M. (2001). *Software Engineering*. Pearson Studium.